ELSEVIER

Virus Research 36 (1995) 97–106

Virus Research

Sequence analysis of putative pVIII, E3 and fibre regions of porcine adenovirus type 3 *

P. Seshidhar Reddy, É. Nagy, J.B. Derbyshire *

Department of Veterinary Microbiology and Immunology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Received 8 August 1994; revised 8 November 1994; accepted 14 November 1994

Abstract

Sequence analysis of a region of the genome of porcine adenovirus type 3 from map unit 79.5 to map unit 92 was performed. Homology studies revealed genes coding for the hexon-associated protein pVIII on the left and for the fibre protein on the right of the sequenced region. By analogy with the genomic organization of other adenoviruses, the 1179 bp sequence between the pVIII and fibre open reading frames, extending from map unit 81.3 to map unit 84.7, was identified as the equivalent of the E3 region of human adenoviruses. The deduced amino acid sequence of one of the three open reading frames of the putative E3 region showed homology with the 13.3K E3 protein of canine adenovirus type 2. The primary structure of the putative fibre protein was similar to that described for human adenovirus types 2 and 5, with a 14 pseudorepeat motif in the shaft region of the fibre. A 742 bp tandem repeat starting in the middle of the fibre gene and extending beyond the termination codon of this gene was observed.

Keywords: Porcine adenovirus; Sequence analysis

Five porcine adenovirus (PAV) serotypes have been identified to date (Derbyshire et al., 1975; Hirahara et al., 1990). There has been recent interest in the potential use of PAVs as viral vaccine vectors in swine (Tuboly et al., 1993). The use of human adenoviruses (HAV) as vectors for foreign genes in several species, including pigs, is well established (Prevec et al., 1989). The HAV genome is organised into complex transcriptional units: the early region genes and late region

* The nucleotide sequence information presented in this paper is deposited with the GenBank database under accession number U10433.

* Corresponding author. Fax: +1 (519) 767-0809.

SSDI 0168-1702(94)00105-7

genes. The early region genes include the E3 region, which was found to be non-essential for viral replication in cell culture, although some of the gene products were shown to be involved in modulating host immune response in vivo (Wold and Gooding, 1991). Foreign genes up to 4 kb have been expressed in HAVs with deletions in the E3 region and these recombinant viruses expressing foreign genes were shown to be stable, and they replicated in cell culture as efficiently as wild type virus (Bett et al., 1993).

We are proposing to use PAV-3 as a vector for foreign genes in swine. As a first step in the development of PAV-3 as a vector, we characterized the genome of PAV-3 by physical mapping (Reddy et al., 1993). The sequence analysis of the E3 region and the neighbouring pVIII and fibre regions of the PAV-3 genome is described in the present paper. During the course of these studies, similar reports on the molecular biology of the NADC-1 isolate of PAV-4 have been published (Kleiboeker et al., 1993; Kleiboeker, 1994).

The 6618 strain of PAV-3 (Clarke et al., 1967) was cultivated in the swine testis (ST) cell line. The cells were grown in Eagle's minimum essential medium supplemented with 10% neonatal calf serum. Viral DNA was extracted from infected cell monolayers by a modified Hirt method described by Reddy et al. (1993). PAV-3 DNA was blunt-ended by alkali treatment (Sira et al., 1987). The *Xba*I B fragment, which is a right terminal fragment extending from map unit 58 to map unit 100, was cloned into the *Xba*I-*Sma*I sites of pGEM-7Zf(+), obtained from Promega. The *Kpn*I-*Bam*HI fragment covering map units 79.5–92.0 (Fig. 1) was sub-cloned into pGEM-7Zf(+). Nested set deletions in both orientations were made using exonuclease III and S1 nuclease as described by Henikoff (1984). The sequence beyond the *Bam*HI site at map unit 92.0 was obtained using the cloned *Bam*HI C fragment. The clones were sequenced using the SP6 and T7 promoter specific primers by the dideoxy nucleotide chain termination technique (Sanger et

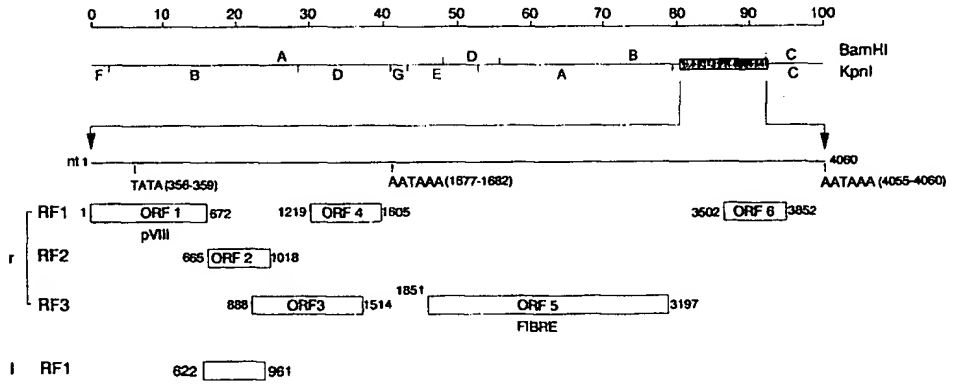

Fig. 1. Physical map of PAV-3 genome for the restriction enzymes *Bam*HI and *Kpn*I. The shaded area (map units 79.5–92) represents the region of the genome which was sequenced. The ORFs in the r- and l-strands with a coding capacity for 100 or more amino acids in the three reading frames, and the locations of TATA and AATAAA signals, are also indicated.

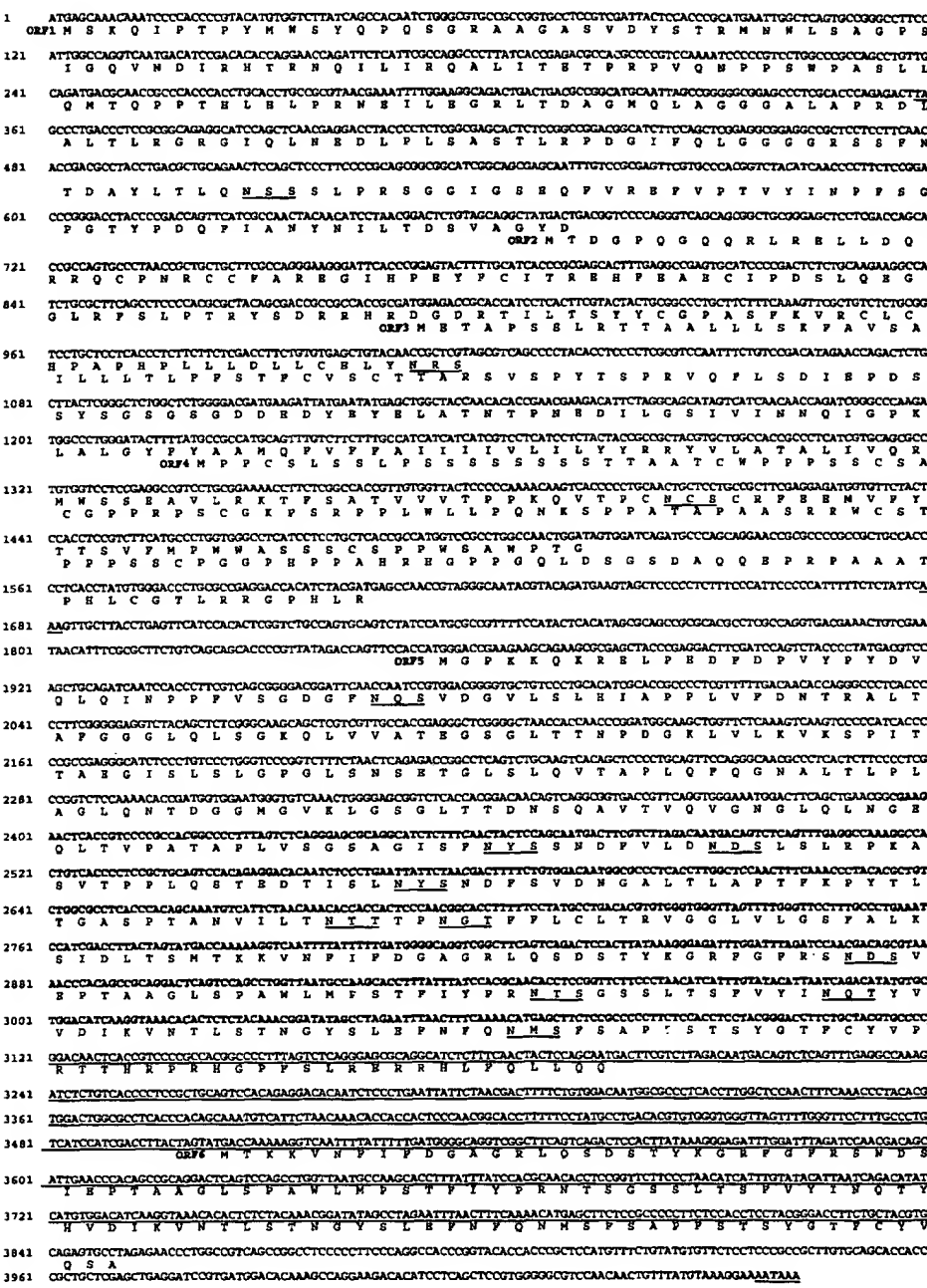

Fig. 2. Nucleotide sequence of the PAV-3 genome between map units 79.5 and 92. The TATA box and two polyadenylation signals (AATAAA) are underlined. The first part of the tandem repeat, starting at nucleotide 2393 in ORF 5, is indicated in bold, and the second part is underlined. The predicted amino acid sequence for the six ORFs on the r-strand with the capacity to encode polypeptides of at least 100 amino acids is indicated below the nucleotide sequence, and potential glycosylation sites (N-X-T/S) are underlined.

al., 1977). A homology search of the GenBank data base was made using BLAST for the nucleotide and deduced amino acid sequences for each of the open reading frames (ORFs). Sequence alignments were carried out using the Palign programme of the PC/GENE sequence analysis software package. Sequence analysis utilized Clone Manager version 3.12 and the Seqaid II version 3.5 programmes.

Physical maps of the PAV-3 genome for the enzymes *Kpn*I and *Bam*HI, the location of the sequenced region in the genome, and of the deduced ORFs in all three frames for the r- and l-strands, are shown in Fig. 1. The 4060 nucleotide sequence, and the predicted amino acid sequence of the six ORFs of the r-strand with a coding capacity of 100 or more amino acids, are given in Fig. 2. The sequence between the *Kpn*I site and the initiation codon of ORF 1 is not shown. The r-strand contained seven ORFs which could encode polypeptides of 50 or more amino acids, while the l-strand included five ORFs. ORFs 1 to 6 on the r-strand (Fig. 1) had theoretical coding capacities for proteins of M_r of 24.2K, 13.7K, 23K, 13.1K, 48K, and 12.8K, respectively. ORFs 1 to 4 were overlapping, whereas ORFs 5 and 6 were non-overlapping. When the 4060 nucleotide sequence was compared with that of other adenoviruses, the only significant homology, in the range of 50 to 60%, was between ORF 1 and the pVIII genes of PAV-4, BAV-3, and HAV types 2, 3, 40 and 41. A TATA box was located within ORF 1, between nucleotides 356 and 359, 316 bp 5′ from the termination of the putative pVIII gene and 309 bp upstream of the ORF 2 ATG. In HAV-2 and CAV-1 a TATA box, located about 320 and 360 bp 5′ of the termination codon of the pVIII gene, was considered to be the promoter for the E3 genes (Linne, 1992). There were two canonical polyadenylation (AATAAA) signals on the r-strand, at nucleotides 1677 and 4055, and four polyadenylation signals on the l-strand, at nucleotides 2796, 2928, 3520 and 3652.

A 724 bp tandem repeat was identified, starting at nucleotide number 2393 in the middle of ORF 5 and extending to nucleotide 3840 in ORF 6 (Fig. 2). While spontaneously occurring adenovirus mutants which contain insertions of viral or host DNA sequences, and deletion mutants, have been reported (Jones and Shenk, 1978), tandem repeats in the adenoviral genome have been described previously only in the ITR of CAV (Sira et al., 1987). The repeat is not a cloning artifact as the same repeat was detected in the genome of PAV-3 by restriction endonuclease analysis (data not shown).

The deduced amino acid sequence of ORF 1 (M_r 24.2K) showed 50% identity plus 13% similarity with the pVIII amino acid sequence of PAV-4, and contained one canonical glycosylation signal (A-X-S/T) at amino acids 169–171. As shown in Fig. 3a, the amino and carboxy terminal regions of the protein showed higher homology than the central region with the pVIII of PAV-4. There is good evidence from the literature that the N- and C-terminal regions of pVIII are more highly conserved among different adenoviruses than the central part of this protein (Raviprakash et al., 1989; Mittal et al., 1992; Kleiboeker, 1994). The highly conserved nature of these regions of the protein emphasizes their functional importance in the structure of the adenoviral capsid.

The deduced amino acid sequence of the protein coded by ORF 2 (M_r 13.7)

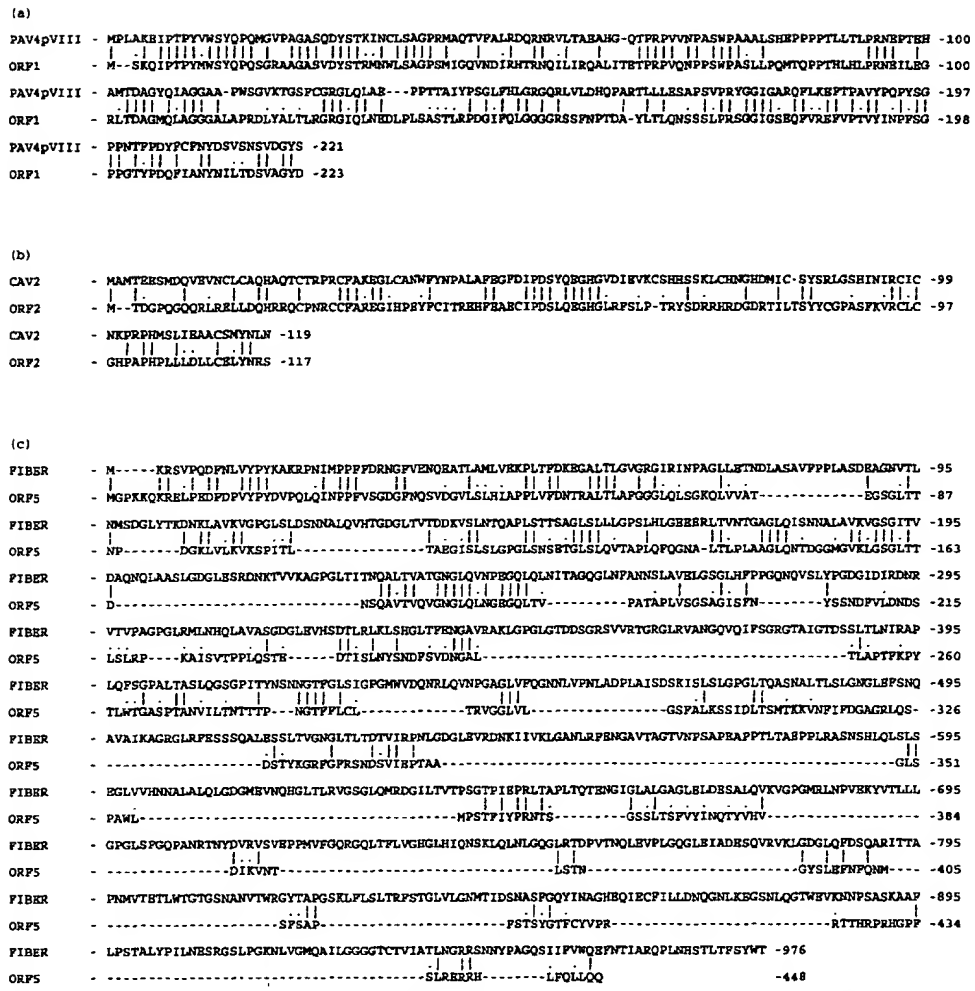

Fig. 3. Comparison between (a) the predicted amino acid sequence of ORF 1 of PAV-3 and the hexon-associated pVIII precursor of PAV-4, (b) the predicted amino acid sequence of ORF 2 of PAV-3 and the 13.3K E3 protein of CAV-2, and (c) the predicted amino acid sequence of ORF 5 of PAV-3 and the fibre protein of BAV-3, using the Palign (PC/GENE) computer program, with comparison matrix structural genetic matrix: open gap cost 6; unit gap cost 2. Identical residues are indicated (|) and similar residues (.).

revealed an identity of 33.3% with the sequence of the 13.3K E3 protein of canine adenovirus type 2 (CAV-2), as shown in Fig. 3b, and contained one potential glycosylation signal at the carboxy terminus of the polypeptide. In CAV-2 and HAV-2 the ORF 1 of the E3 region is followed by a putative L4 polyadenylation signal (Linne, 1992), but no such signal was found in the E3 region of PAV-3.

The GenBank search revealed no homology between the deduced amino acid sequence of ORF 3 and any known adenoviral protein. The amino acid sequence

had one potential glycosylation site, and two membrane-associated helices predicted by the method of Eisenberg et al. (1984). The first helix was located between residues 12 and 32 (AALLLSKFAVSAAILLLTLFF), and was classified as globular, while the second was a trans-membrane multimeric helix located between residues 107 and 127 (LGYFYAAMQFVFFAIIIIVLI). The globular hydrophobic helix could act as a signal sequence, although it is located 12 residues away from the amino terminus. A signal sequence present in the 19K protein encoded in the E3 region of HAV-2 is close to the amino terminus (Persson et al., 1980). Although the predicted amino acid sequence of the polypeptide encoded by ORF 3 did not show homology with the 19K E3 protein of HAV-2, the presence of this putative signal sequence and a transmembrane domain suggests that this protein may resemble the 19K protein of HAV-2 and HAV-5, which binds to the MHC class I antigens in the endoplasmic reticulum and prevents their transport to the cell surface (Wold and Gooding, 1991). ORF 4, with the coding capacity for a protein of M_r 13.1K, did not show homology with any known adenoviral protein. It included a stretch of eight serine residues between amino acids 11 and 18 near its amino terminus.

The deduced amino acid sequence of ORF 5 (M_r 48K) showed 30% identity plus 16% similarity with the fibre protein of BAV-3, and, as shown in Fig. 3c, the N-terminal portion of the protein showed the highest homology (39% identity). The protein sequence of ORF 5 could be divided into tail, shaft and knob components (Fig. 4), and the shaft component was found to be arranged in 14 pseudorepeat motifs, each 15 residues long, in accordance with the model proposed by Green et al. (1983) for HAV-2 and Kidd et al. (1993) for HAV-40. As shown in Fig. 4, the 15 residue pseudorepeat motifs consisted of periodically occurring prolines, glycines and hydrophobic residues, indicating that the relative hydrophobicities rather than strict identities are conserved in the pseudorepeats. However, since there are fewer repeats in the shaft region of the putative fibre in PAV-3 than in the fibres of HAV-2 and BAV-3 (Mittal et al., 1992), the PAV-3 fibre would be expected to be shorter than that of HAV-2 and BAV-3. A short fibre, with only six repeats in the shaft region was also found in HAV-3 by Signas et al. (1985). It was reported by Devaux et al. (1987) that the 16 N-terminal amino acids of the fibre protein of HAV-2 are required for binding of the fibre to the penton base and are conserved among different adenoviruses. Part of such a conserved sequence, PVYPYD, was also present in the tail region of the putative fibre of PAV-3. Although the fibres of HAV-2 and HAV-12 are glycoproteins (Ishbashi and Maizel, 1974; Bruggeman et al., 1985) there is no evidence for N-linked glycosylation in any adenoviral fibre. The deduced amino acid sequence of ORF 5 (Fig. 2) revealed 10 potential glycosylation signals. Most of the adenoviral late mRNAs contain a tripartite leader. In HAV-2 the mRNAs coding for the fibre protein contain three leader sequences in addition to the tripartite leader (Uhlen et al., 1982). No such sequences were found in PAV-3. The polyadenylation signal found at nucleotide 4055 in the present study is most probably used for polyadenylation of the putative fibre mRNA encoded by ORF 5, and perhaps for ORF 6. It is likely that the potential poly A sites located on the l-strand are used

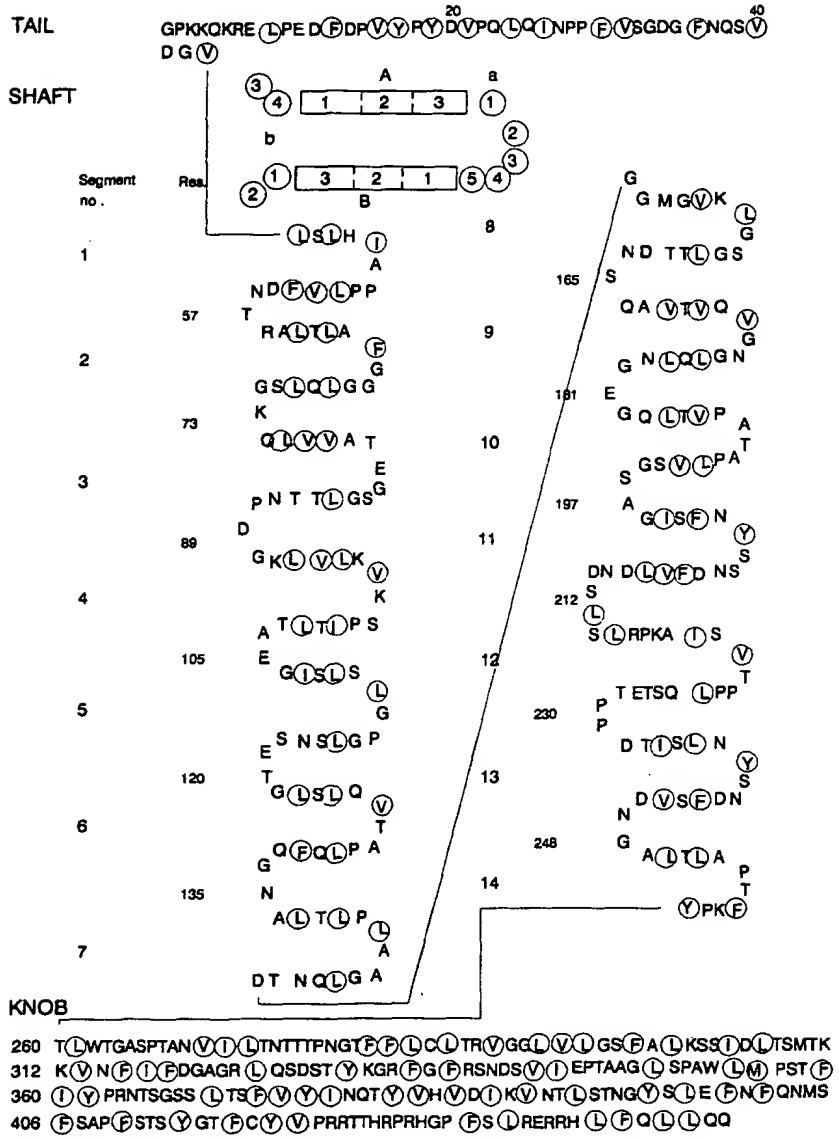

Fig. 4. Structural organization of the subunit of the fibre protein of PAV-3 based on the model proposed by Green et al. (1983) for the fibre protein of HAV-2. Each subunit consists of a short N-terminal tail, 14 repetitive segments which form the shaft, and a distal knob. The repeats in the shaft have two short β strands, A and B, having three amino acid residues each, which are linked to each other by two β bends, a and b, containing 5 and 4 amino acids, respectively (a1 to a5; b1 to b4). Hydrophobic amino acids are circled.

for the poly A tail addition of E4 mRNAs. The deduced amino acid sequences of ORF 6 and of the ORF present on the l-strand showed no homology with any other protein.

In human (Cladaras and Wold, 1985), bovine (Mittal et al., 1992), canine (Linne, 1992) and murine (Raviprakash et al., 1989) adenoviruses the E3 region lies between the genes coding for the precursor protein pVIII and the fibre protein of the L4 and L5 regions of the genome, respectively. The same genomic organization was reported for the NADC-1 isolate of PAV-4 by Kleiboeker (1994). Since in the case of PAV-3 the predicted amino acid sequence of ORF 1 revealed homology with the pVIII gene of PAV-4, and the deduced amino acid sequence of ORF 5 showed 30% identity with the BAV-3 fibre protein, it seems likely that ORF 1 and ORF 5 are the genes coding respectively for the precursor pVIII and fibre proteins of PAV-3. Based on these observations it is assumed that the E3 region of PAV-3 is located between ORF 1 and ORF 5. The E3 region of PAV-3 would therefore be 1179 bp long and be located between map units 81.3 and 84.7. In HAV-2 and HAV-5 the length of the E3 region is about 3 kb, but shorter E3 regions have been reported for several adenoviruses of other animals: 782 bp in MAV-1 (Raviprakash et al., 1989), 1.1 kb in CAV-1 (Dragulev et al., 1991), 1.5 kb in BAV-3 (Mittal et al., 1992) and 1.0 kb in PAV-4 (Kleiboeker, 1994). The E3 region of animal adenoviruses seems to be somewhat variable in size, but consistently smaller than the E3 region of HAV genomes.

In HAVs, the E3 region is complex and codes for several overlapping mRNAs. The number and type of products vary from serotype to serotype. The E3 regions of HAV-2 and HAV-5 include ten ORFs (Cladaras et al., 1985), while the number of ORFs encoding polypeptides of 50 or more amino acids in the E3 region of other animal adenoviruses varies from one in MAV-1 (Raviprakash et al., 1989), two in CAV-1 (Dragulev et al., 1991) to three in BAV-3 (Mittal et al., 1992) and PAV-4 (Kleiboeker, 1994). The present study has demonstrated three overlapping ORFs in the E3 region of PAV-3. In HAV-2 and HAV-5, transcripts of the E3 region have been divided into two 3′ co-termination families depending on which polyadenylation signal is used (Cladaras et al., 1985). In PAV-4 and CAV-1 only one polyadenylation signal has been reported (Dragulev et al., 1991; Kleiboeker, 1994), and in the present study of PAV-3 only one polyadenylation signal (AATAAA), most likely used for the polyadenylation of E3 mRNAs, was located 72 nucleotides from the end of ORF 4. Based on these observations it can be concluded that the E3 region of PAV-3, as in other animal adenoviruses, has fewer ORFs and is much simpler in organisation than the E3 region of HAVs.

As the predicted amino acid sequence of the polypeptides encoded by ORF 4 and ORF 6 did not show homology with any known adenoviral protein, the functional significance of these two ORFs is unknown. The highly variable nature of products of the E3 region in different adenoviruses compared to the highly conserved flanking L4 and L5 regions suggests that the E3 region is most probably used by adenoviruses to acquire new genes which might not be essential for viral replication in vitro but might be required for the long-term survival of virus in nature.

Among the PAVs, sequence data comparable to those reported in the present paper for PAV-3 are available only for PAV-4 (Kleiboeker, 1994). While the genomic organization of the two viruses appeared to be similar in the sequenced

region, the putative E3 region of PAV-3 was shorter than that of PAV-4, but it contained the same number of ORFs, and a single polyadenylation signal. Except for ORF 1, encoding the putative pVIII protein, which showed 60% homology with the corresponding ORF of PAV-4, none of the PAV-3 ORFs showed significant nucleotide sequence homology with those of PAV-4. The deduced amino acid sequence of ORF 1 and the first 48 amino acids at the amino terminus of ORF 5 of PAV-3 showed identities of 50% and 48%, respectively, with the corresponding pVIII and fibre proteins of PAV-4. The deduced amino acid sequences of the putative E3 ORFs of PAV-3 showed no significant homology with those of PAV-4.

Acknowledgements

The research was supported by the Natural Sciences and Engineering Research Council of Canada, and by the Ontario Ministry of Agriculture, Food and Rural Affairs. We acknowledge the technical assistance of John Dennis and Neeraja Idamakanti.

References

Bett, A.J., Prevec, L. and Graham, F.L. (1993) Packaging capacity and stability of human adenovirus type 5 vectors. J. Virol. 67, 5911–5921.

Bruggeman, U., Klenk, H.-D. and Doerfler, W. (1985) Increased infectivity of extracelluar adenovirus type 12. J. Virol. 55, 117–125.

Cladaras, C. and Wold, W.S.M. (1985) The sequence of the early E3 transcription unit of adenovirus 5. Virology 140, 28–43.

Cladaras, C., Bhat, B. and Wold, W.S.M. (1985) Mapping the 5′ ends, 3′ ends and splice sites of mRNAs from the early E3 transcription unit of adenovirus 5. Virology 140, 44–54.

Clarke, M.C., Sharpe, H.B.A. and Derbyshire, J.B. (1967) Some characteristics of three porcine adenoviruses. Arch. Ges. Virusforsch. 21, 91–97.

Derbyshire, J.B., Clarke, M.C. and Collins, A.P. (1975) Serological and pathogenicity studies with some unclassified porcine adenoviruses. J. Comp. Pathol. 85, 437–443.

Devaux, C., Caillet-Boudin, M.-L., Jacrot, B. and Boulanger, P. (1987) Crystallisation, enzymatic cleavage and the polarity of the adenovirus type 2 fiber. Virology 161, 121–128.

Dragulev, B.P., Sira, S., Abouhaidar, M.G. and Campbell, J.B. (1991) Sequence analysis of putative E3 and fiber genomic regions of two strains of canine adenovirus type 1. Virology 83, 298–305.

Eisenberg, D., Schwartz, E., Komarony, M. and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179, 125–142.

Green, N.M., Wrigley, N.G., Russel, W.C., Martin, S.R. and McLachlan, A.D. (1983) Evidence for a repeating cross-β sheet structure in the adenovirus fibre. EMBO J. 2, 1357–1365.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. Gene 28, 351–359.

Hirahara, T., Yasuhara, H., Matsui, O., Yamanaka, M., Tanaka, M., Fukuyama, S., Izumida, A., Yoshiki, K., Kodama, K., Nakai, M. and Sasaki, N. (1990) Isolation of porcine adenovirus from the respiratory tract of pigs in Japan. Jpn. J. Vet. Sci. 52, 407–409.

Ishbashi, M. and Maizel, J.V. (1974) The polypeptides of adenoviruses VI. Early and late glycopolypeptides. Virology 58, 345–361.

Jones, N. and Shenk, T. (1978) Isolation of deletion and substitution mutants of adenovirus type 5. Cell 13, 81–188.

Kidd, A.H., Chroboczek, J., Cusack, J. and Ruigrok, R.W. (1993) Adenovirus type 40 virions contain two distinct fibers. Virology 192, 73–84.

Kleiboeker, S.B. (1994) Sequence analysis of putative E3, pVIII and fiber genomic regions of a porcine adenovirus. Virus Res. 31, 17–25.

Kleiboeker, S.B., Seal, B.S. and Mengeling, W.L. (1993) Genomic cloning and restriction site mapping of a porcine adenovirus isolate: demonstration of genomic stability in porcine adenovirus. Arch. Virol. 133, 357–368.

Linne, T. (1992) Differences in the E3 region of the canine adenovirus type 1 and type 2. Virus Res. 23, 119–133.

Mittal, S.K., Prevec, L., Babiuk, L.A. and Graham, F.L. (1992) Sequence analysis of bovine adenovirus type 3 early region 3 and fibre protein genes. J. Gen. Virol. 73, 3295–3300.

Persson, H., Jornvall, H. and Zabielski, J. (1980) Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: identification and structure of the signal sequence. Proc. Natl. Acad. Sci. U.S.A. 77, 6349–6353.

Prevec, L., Schneider, M., Rosenthal, K.L., Belbeck, L.W., Derbyshire, J.B. and Graham, F.L. (1989) Use of human adenovirus-based vectors for antigen expression in animals. J. Gen. Virol. 70, 429–434.

Raviprakash, K.S., Grunhaus, A., El Kholy, M.A. and Horwitz, M.S. (1989) The mouse adenovirus type 1 contains an unusual E3 region. J. Virol. 63, 5455–5458.

Reddy, P.S., Nagy, É. and Derbyshire, J.B. (1993) Restriction endonuclease analysis and molecular cloning of porcine adenovirus type 3. Intervirology 36, 161–168.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Signas, C., Akusjarvi, G. and Pettersson, U. (1985) Adenovirus 3 fiber polypeptide gene: Implications for the structure of the fiber protein. J. Virol. 53, 672–678.

Sira, S., Abouhaidar, M.G., Liu, Y.C. and Campbell, J.B. (1987) Multiple reiteration of a 40 bp nucleotide sequence in the inverted terminal repeat of a canine adenovirus. Virology 159, 76–83.

Tuboly, T., Nagy, É. and Derbyshire, J.B. (1993) Potential viral vectors for the stimulation of mucosal antibody responses against enteric viral antigens in pigs. Res. Vet. Sci. 54, 345–350.

Uhlen, M., Svensson, C., Josephson, S., Alestrom, P., Chattapadhyaya, J.B., Pettersson, U. and Philipson, L. (1982) Leader arrangement in the adenovirus fiber mRNA. EMBO J. 1, 249–254.

Wold, W.S.M. and Gooding, L.R. (1991) Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus cell interactions. Virology 184, 1–8.